DOCUMENT RESUME

ED 459 761 IR 058 255

AUTHOR	Nichols, Stephen G.
TITLE	Artifacts in Digital Collections.
PUB DATE	2001-08-00
NOTE	9p.; In: Libraries and Librarians: Making a Difference in the Knowledge Age. Council and General Conference: Conference Programme and Proceedings (67th, Boston, MA, August 16-25, 2001); see IR 058 199.
AVAILABLE FROM	For full text: http://www.ifla.org.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Access to Information; *Electronic Publishing; Evaluation Criteria; Information Technology; *Library Materials; Library Role; *Library Services; *Preservation
IDENTIFIERS	Artifacts; Digital Data; *Electronic Resources; *Information Value

ABSTRACT

This paper discusses the preservation of digital resources in research libraries. The first section addresses the preservation imperative, including the challenges of quantity, stability of media, economics, and the contingent value of artifacts. The second section describes the artifact in question (i.e., an information resource in which the information is recorded on a physical medium, which may or may not be unique, and in which the information value of the resource adheres not only in the text or content, but also in the object itself). Topics covered in this section include definition of the term "artifact," selection for the preservation of the original format, and mechanisms for determining value. (MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 459 761

67th IFLA Council and General Conference August 16-25, 2001

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

A.L. Van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Code Number: 084-168-E
Division Number: VI
Professional Group: Preservation and Conservation
Joint Meeting with: Information Technology
Meeting Number: 168
Simultaneous Interpretation: -

1

Artifacts in digital collections

Stephen G. Nichols

Romance Languages Department, Johns Hopkins University
Chair; Task force on the Artifact in Library Collections
Baltimore, USA
E-mail: stephen.nichols@jhu.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

1. The Preservation Imperative

The "information explosion" sparked by digital technology has fostered an increasing awareness of the sheer mass of information available today in a variety of media, from traditional formats like paper to the more recent film, optical, and magnetic formats. Today, the world produces two billion gigabytes of new information a year, or roughly 250 megabytes for every man, woman, and child on earth (Berkeley 2000). Institutions charged with collecting, storing, preserving, and making accessible recorded information have struggled to keep pace with the growth of information production, even though their brief is to collect only a portion of what is published and an even smaller portion of what is produced and disseminated in unpublished form.

Although information overload is not a new problem, the introduction of digital technology onto campuses and into research libraries has fundamentally altered the information landscape, with potentially serious ramifications for scholars and students. The creation and dissemination of digital resources are creating new models of service and access, such as licensing rather than owning essential intellectual assets. The mutability of digital documents is redefining what constitutes a text—are back issues of a journal that are in digital form a bunch of articles or a rich database? And the conversion of texts into hypertexts is resulting in increased interdisciplinary research, when a researcher in one field serendipitously find resources normally confined (in print) to another.

IR058255

BEST COPY AVAILABLE

Accompanying these trends is another series of trends that at first seem paradoxical. There is increased scholarly attention to original, unreformatted materials. And there are eruptions of public outcry when discoveries of material losses in libraries and archives are made. While scholars demand ever-increasing attention to an ever-expanding range of candidates for preservation, library budgets simply do not support those demands. Preservation has thus become an "unfunded mandate," the more pernicious for being often implicit. Academic institutions have learned the huge costs of penny-wise facilities management and "deferred maintenance." It is reasonable to fear that we are now incurring similar future costs by deferring preservation. With preservation, though, there is a crucial difference: a significant part of what we avoid thinking about today will be lost forever by our neglect. Library collections are among the most valued of a research institution's intellectual and cultural assets, assets that form a crucial part of what might be called "public goods." But, as is the case with public goods, many make claim to their use, but few the responsibility for their well being. This paper focuses on preservation—what it takes to ensure the present and future usability of collections—for without preservation today, there will be no access tomorrow.

Preservation is a critical part of good stewardship of our intellectual and cultural heritage. Its chief challenges at the turn of the twenty-first century are fourfold:

Quantity. Because of the relentless growth of research libraries and their collections, an immense number of research items demand resources to remain accessible. In 1999, the 121 member libraries of the Association of Research Libraries (ARL) reported owning a combined total of 462,965,204 volumes. The greatest period of growth for research libraries occurred after the First World War, after the Second World War, and in the last decade. Looking at 12 representative public university libraries during this period reveals a typical growth pattern.[1] In 1907, these libraries held an average of 107,425 volumes; by 1961, that number had grown to 1,772,831, and by 1995, the average per library was 5,334,620 (Gerould).

Stability of media. Library collections exist in a variety of physical formats that are vulnerable to one degree or another. As the rate of information production has increased, the storage medium has become more compact and efficient. However, this miniaturization comes as the expense of stability and longevity. With the exception of preservation-quality microfilm, the new media of the twentieth-century are more fragile than those of the nineteenth, including the by-now infamous wood-pulp paper that deteriorates into flakes over time (Conway 1996, 5). The many media that have been invented in the last 150 years to capture light and sound are generally extremely fragile, dependent on machines for playback, and subject to rapid technological obsolescence. The wax cylinders on which are inscribed the earliest known recorded voices of Native Americans are susceptible to mold, heat, scratching, skin oils, and physical trauma. Moreover, they are dependent on playback equipment for which no spare parts are manufactured and for which there are few technicians skilled in repair. Yet the information on them is invaluable and must be saved for future generations.

Economics. Since 1993, preservation budgets in ARL libraries have remained flat and, in real dollars, are shrinking. At the same time, the number of staff assigned to preservation is at a ten-year low (Scott 1999). Yet the demand for access to original materials has grown, especially to those materials in special formats that often are at greatest risk from physical handling or environmental stress. Meanwhile, as the technology for reformatting for access has greatly improved, the funding for preservation continues to decrease, since more money goes to digital reformatting of items in high demand and less to microfilming the low-use brittle books that are rotting on shelves.

---

[1] The libraries are: University of California, Berkeley; University of Illinois; Indiana University; University of Iowa; University of Kansas; University of Michigan; University of Minnesota; University of Missouri; University of Nebraska; Ohio State University; University of Washington (Seattle); and the University of Wisconsin.

Contingent Value of artifacts. The most difficult challenge for libraries is deciding how to set priorities for preservation. As long as the claim on preservation resources exceeds the available funds, from both internal and external sources, there will be a necessary triage of materials that get treatment. An example of this trend is the recent elevation of nineteenth-century popular imprints and ephemera to a status of high research value. Providing appropriate access to those items, which are at high risk from embrittlement and ordinary physical handling, has put an enormous strain on library resources. Knowing that the intellectual interests and research methodologies of scholarship will change over time, research institutions have collected "just in case" there is a demand in the future, rather than "just in time" for demand. Research libraries and archives are full of items that have not been consulted in decades, if ever, and whose future demand is unpredictable. How do a library and its home institution make sound fiscal and intellectual decisions about what to preserve, when, for whom, and at what price? Despite the enormous collections of printed materials that have been amassed, entire categories of primary sources have disappeared before collecting institutions and their users understood their value. The most notorious example of neglect is the fact that 80 percent of all silent films made in the United States are gone without a trace, and 50 percent of films made before the Second World War have also perished and will never be recovered. The loss is so great that we do not even know what has been lost.

2. The Artifact in Question

"Artifact" is a term that can be confusing because it masks a number of unexamined assumptions. In common academic parlance, "artifact" can be used to mean a physical object, a primary record, or a physical object that constitutes a primary record.[2] From the point of view of a researcher, however, and for the purposes of this report, an artifact can be defined as an information resource in which the information is recorded on a physical medium (such as a photograph or a book), which may or may not be unique (holograph letter or a paperback book), and in which the information value of the resource adheres not only in the text or content, but also in the object itself. In other words, artifacts are things that have intrinsic value, independent of the informational content.

Recently, scholars have identified an increasing number of library items that have intrinsic research value as physical objects. The Modern Language Association (MLA) defines an artifact or primary record as "a physical object produced or used at the particular past time that one is concerned with in a given instance" (MLA 1996). The definition goes on to assert that, for all practical purposes, all historical publications, even those produced in mass quantities by mass production techniques designed to minimize deviations from a norm, may have some qualities of uniqueness.[3] Since uniqueness is one of the defining features of something that has artifactual value, this position would seem valid. The problem for libraries, however, is that by definition these items would therefore have a claim not only on the attention of the scholar, but also on the budgets of preservation departments. Whether one disagrees or accepts in toto this definition and its implications, the difficulty it presents for libraries is that they are not, have not been, and are unlikely ever to be funded to meet the need to collect and preserve literally everything of potential research value. For libraries, this expansive view of artifactual value presents problems that are not primarily theoretical, but eminently practical.

Given the task of identifying achievable, fundable preservation strategies and goals for libraries, we must identify intrinsic artifactual value and to do so in way that, following the spirit of preservation principles, would accept some loss as inevitable and sought rather to manage the risks of *unacceptable* loss.

---

[2] Of course, in scientific laboratories, "artifact" also denotes a substance that is a by-product of some external action or agent.

[3] The statement addresses only text-based sources. If this standard of value were extended to visual and sound resources, the universe of primary records grows exponentially.

2.1 Selection for Preservation of Original

The library preservation community has agreed on certain cardinal features of physical objects that warrant preservation in their original formats. They are:

Age
Evidential value
Aesthetic value
Scarcity
Associational value
Market value
Exhibition value

Objective criteria or established practice determine many of these features, and the criteria vary little among libraries. They are, in short, best practice; several written policies are included in Appendix III as meaningful examples.

We are not talking here about those categories of artifacts that are always, as a matter of course, retained in the original. That would mean books printed before 1801, which are usually segregated from general holdings in a rare book collection and subject to different handling and preservation protocols; manuscripts and archival materials that exists only in one copy; items with high market value, and so on.

The value of the artifact for research purposes—as opposed to monetary value or exhibition value—can be thought of as chiefly evidentiary. An artifact is of value to the extent that it testifies to the information being original, faithful, fixed, or stable.

Originality: An original manifestation of a book, photograph, or recorded performance is valuable because through it a scholar may come closer to uncovering the original intent of the creator and or publisher, or both.[4] When copies yield insufficient information about original intent, then access to the original may be needed. Reformatting and copying information is analogous to translation from one language to another. Depending on the source and the target language, as well as the skill, care, and cultural biases of the translator, something is always lost. A good translation, like good recopying, is one that loses the least amount of original content and intent.

Fidelity: The artifact is also useful and at times essential in establishing the authenticity of an item. In other words, it has forensic value. How does one know that the item in one's hands is what it purports to be? There are internal clues in a document that give evidence of authenticity, of course, such as the accuracy and appropriateness of the content. A diary that is dated 1901 and contains references to television broadcasts, for example, is unlikely to be what it purports to be. But in addition to the intellectual information, there is also the external information contained in a physical manifestation that also provides clues of authenticity and integrity. Erasure marks on a sheet of paper, splices in a film, white-out on property maps—all these are physical clues to the integrity of the object and, hence, the authenticity of the information recorded in the object.

---

[4] The instances of published versions differing from the presumed intent of the creator are legion in books, films, and so forth. In those cases, the sources that contain information about the work in pre-publication form (e.g., draft, outtakes, proof sheets) are also required to reconstruct original intent.

Fixity: The content of the artifact when it was first produced constitutes the text (in the case of textual materials) or the document (in the case of a photograph, say, or an opera performance). If one is holding a fifteenth-generation fax, one cannot guarantee that the full content of the original is conveyed except by comparison with the original, which has fixed the content by recording it at one instant in time. One of the wonders of mechanical reproduction of text after Gutenberg was the way that replication by machine en masse tended to fix texts that had previously been taken to be somewhat fluid. This is one of the innovations of the print regime that appears to be eroding in the digital realm.

Stability: The persistence of an object over time leads to the stable and continuous accessibility of the information contained in it. Documents whose physical substrate changes over time themselves change. Film that gets spliced and repaired loses content; digital files that get reformatted into a newer version of a software program change; photographic images printed or displayed in various manner shift tone. When one looks at a thirty-year-old image of a woman in a red coat printed on paper that fades, and looks at a contemporary image made from the same negative projected through a slide projector, chances are there will be two different reds that constitute "the coat." The content of that red is not stable, and it is difficult to mentally efface the effects of age and reformatting and determine whether the coat was originally scarlet red or crimson red.

There are, in addition, artifacts valuable for research because the format is itself the subject of investigation. Original bindings carry evidence of print history, just as original daguerreotypes carry evidence of an early photographic technology. In these cases the object itself is the primary source, not the information it carries.

Also of value to the research process is the physical encounter between the researcher and the object, an encounter that can help to prime the scholars' imaginative and analytical skills. While this factor is highly subjective and difficult to quantify, it is commonly cited by scholars as being, at least at some stage of their career, of irreplaceable heuristic value. A medievalist who has never worked directly from manuscripts is at a disadvantage, just as a biographer of Thomas Jefferson who has worked exclusively from the printed editions of his letters may be said to work at one critical remove from the subject. Given the toll that physical handling takes on all types of materials, though, scholars and library professionals should accept the fact that surrogates can be judiciously used by those who have a familiarity with original source materials and, from the perspective of both preservation and convenience of access, are often preferable. [5]

2.2 Mechanisms for Determining Value

Questions about the nature of the artifact have caused scholars and library professionals to realize that, even for the early part of the nineteenth century, much more information of potential research value exists in traditional formats, such as paper and image, than had previously been recognized. Consequently, the process of redefining what constitutes an artifact must be done not only for new media, but also for a considerable body of information from the 1800s. The fragility of paper-based materials, especially newspapers, printed since 1850 has been a concern for some time (Marley, Baker, Smith, Cox). Because of their fragility, preserving one or more instances of all imprints of newspapers poses enormous technical as well as financial challenges to collecting institutions. More recently, there has been a growing awareness of other kinds of artifacts from the nineteenth and early twentieth centuries that also require the attention of

[5] See, for example, the case of the editorial team working on the James Boswell Papers at the Beinecke Library (Bouché 1999). The international team of editors came to prefer the use of digital scans of the original manuscripts to working from the originals in New Haven. In part this was the matter of convenience—the work could be done wherever the editors were located and obviated a series of disruptive trips to New Haven—and in part because the scans provided superior legibility.

preservationists if they are not to disappear. Often, these are materials that serve as primary evidence of popular culture that were neglected by many researchers until recently. Such materials range from baseball cards to ladies' magazines to dime novels to political posters—all matter produced for a mass market and using the cheapest and most available materials.

The examples from popular culture raise perhaps that most difficult of questions: how to weigh the relative merits of various claims on scarce preservation resources?

First, let us distinguish the issues. Scholars, archivists, and librarians have always assumed a hierarchy in collections. The artifact or original document was the item initially collected and accessioned in the collection. When it was prized for some unique aspect of its material or historical existence (e.g., status as a first edition, holograph manuscript, signed author's copy or presentation copy of a work) the artifact was given priority for preservation. In the absence of that uniqueness, a lower level of care might be given and a lost or damaged copy replaced with one very different in physical manifestation (e.g., a photocopy or reprint). The value of the unique artifact could be defined variously as historically important, rare, associational, and so on. In each such instance, however, there was an individual, material object that someone, somewhere defined as valuable enough to retain.

The Achilles heel of traditional definitions of the artifact lay in the value judgment that determined artifactual status in the first instance. What were the grounds for deciding in favor of one object and against another? How can libraries cope with the fact that the value of the artifact is never quite the same to different researchers? While one scholar will seek certain information from the item, others from different disciplines will require other kinds of data, perhaps involving a wholly different way of handling the object. [6] Pragmatically, can one say that these users are even seeing the same object?[7]

Scholars and others who use artifacts may think of them in the aggregate as unified objects. On the other hand, when they come to interrogate the artifact for their research, they tend to focus on parts rather than the whole (which is not, of course, to say that the relationship of part to whole is not crucially important). One may define the artifact as a series of multiple discrete components—handwriting, watermarks, marginalia, splices, evidence of use—each potentially a focal point for scholars and others, depending on what they are studying. From the standpoint of usage, what people normally do with artifacts is to analyze discrete sets of information contained within the object. The fact that artifacts are complex, that they lend themselves to a variety of intellectual endeavors, means that we must think of them in terms of their parts, and not just as a whole. One way to think of the artifact, then, would be as a multiplicity of informational sets, including the material form of the object, and its contextual history, where known.

Not the least aspect of the artifact's complexity is the fact that different scholars from different fields will perceive and use it in different ways. A first edition of a novel by Dickens will be used very differently by an historian of Victorian England researching the economics of the book trade, a literary scholar concerned with different versions of the work, an art historian interested in Victorian book illustration, a textual

---

[6] Research disciplines train scholars to attend to the materiality of their objects of study in very different ways. It is not a criticism, but a fact of research, that historians, literary scholars, philosophers, art historians, historians of science, linguists, text editors, and so on, are *supposed* to examine quite different aspects of an object of study, which, in the case of artifactual objects, will certainly be complex. Sometimes there may be overlap, more often not. The needs of all serious users are legitimate, and preservation should serve them as effectively as possible.

[7] Artifacts derive their value from how they are viewed and used in a given culture at a particular moment. As cultural variables, they will be viewed and studied differently in different periods. A holograph copy of a speech by Robespierre would certainly have a different value for a royalist in Louis XVIII's government in 1816 then it would a socialist historian in France in the 1990s.

historian interested in layout, and an historian of book making. Each will consult the same artifact for very different kinds of information, and each one may well not even remark (or have the technical knowledge to do so), the particular information sets sought by the other researchers. Similarly, photographs of the Civil War by Alexander Gardner can be used to study the battles, or the public's reception of the war in North, of the history of clothing, of medicine, of gender, or of the medium itself. Which subject interest and methodology would require use of the original, which could make do with copy prints, which could make use only of the original photographs in their original presentation portfolios, and which would actually be enhanced by access to the images through digital delivery which could then be manipulated to magnify details, for example?

The preceding observations suggest the possibility of proposing a contextual definition of the artifact as follows:

> *An artifact is a physical object produced at some time in the past, and attesting to a given set of practices, thinking, and ways of viewing the world, but whose importance will be defined by present and foreseeable future needs and use. The value of the artifact is strongly influenced, but not completely determined by, the document/object's features that are unique.*

The artifact conveys historical consciousness in a number of different ways, depending on who studies it and for what purpose. Much of the information conveyed by an artifact does not require the presence of the physical object. Surrogates of the object (e.g., photographs, photocopies, and digital versions) may convey much of the information stored by the artifact. For some, perhaps many, purposes a high-quality surrogate may convey this information better than does the original. In many instances, the surrogate may enable access and use that would be impossible if we were dependent on the original. Access to objects that are remote and the ability to enhance images or perform full-text searches are among the uses that can render some difficult-to-use materials suddenly accessible. Surrogates do not obviate the need for some scholars to consult the object itself. However, there are many instances when a surrogate can reasonably serve scholarly needs as well as, sometimes better than, the artifact itself.

This is all by way of explaining that the artifact matters. It matters very much. Yet faced with the crisis in which artifacts are abundant and resources scarce, the scholarly and library communities urgently need to rethink the status of the artifact in terms of its content and its material form. Acknowledging that every aspect of an artifact yields information that will be of use to some scholar or other, we may need to assess the relative importance of the different aspects of an artifact pragmatically. Several issues need to be confronted collectively by scholars and library professionals.

First, within the timeframe of the last two hundred years, what constitutes an artifact worth retaining? The answer is not obvious in the case of nineteenth-century material artifacts: baseball programs or railway timetables have not traditionally tended to be viewed as important cultural documents, although they probably would be today. However, it becomes truly perplexing in the case of the electronic medium, where the concepts of original or unique, stable or fixed, may not even apply.

A second question concerns resources and priorities. This imperative may be spelled out as the "how, who, and when" of artifact preservation. Accepting the reality that resources for preservation are limited, and assuming some consensus on the nature of artifacts: how (in what form) are they to be preserved, by whom, and when (or how often)? In other words, are all libraries or archives to be held responsible for collecting and preserving the same categories of artifacts? How much redundancy of preservation is necessary and how much can we afford? The presence of redundant collections is an insurance policy for preserving and making accessible information in physical formats. However, this is untrue for digital information, which

can be easily cloned and shipped around the globe almost instantaneously. In the case of digital information, redundancy may be very wasteful of scarce resources. If and when digital files become the default mode for access—even for materials that were originally physical artifacts--what are the implications for duplicative collections of physical artifacts—old runs of journals, embrittled monographs, and so on?

The issue is less to determine whether baseball programs and railway timetables from the nineteenth century, for example, are items of sufficient artifactual value to justify the expense of collecting and preserving them. It is rather one of setting priorities in the face of financial constraints that mean, practically speaking, that if one decides to collect and preserve one kind of artifact, there will be insufficient resources in a given library or archive for other kinds.

Nor is the issue to evaluate the artifact per se to determine what survives and what does not—the scholarly community has no more of a claim to the wisdom of the ages than does the library community. What is needed is to agree on a system, a method for interrogating the individual artifact that would, in a climate of finite resources, help make a good decision about whether and how to preserve it. Such a system would help assure survival of the greatest number of artifacts by intelligent analysis and classifications.

If the scholarly and library communities could reach consensus about matching preservation of specific kinds of artifacts to particular categories or kinds of repositories, it will be possible to rationalize on a national scale the task of artifact preservation.

There are few effective mechanisms available for creating guidelines or determining where the responsibilities fall for making decisions about retention, preservation, and the kinds of reformatting that might be deemed appropriate for a given artifact type. There have been initiatives that have brought together the scholarly, archival, and libraries communities to propose preservation and collection guidelines. The Commission on Preservation and Access convened scholars in the 1980s to identify brittle books to microfilm. More recently, in the 1990s, the American Historical Association, the Association of Research Libraries, and the Modern Language Association created a task force on the preservation of the artifact to document the preservation challenges and inform scholars about the need to collaborate with libraries to address those problems.[8] However, many of the collaborations between scholars and librarians have been either locally based or ad hoc, to address specific crises. None of the initiatives has proven to be sustainable, and none has developed mechanisms for long-term productive partnerships.

In 1999, the Council on Library Information Resources, with the encouragent of the American Council of Learned Societies and several foundations, created the Task Force on the Artifact in Library Collections to address precisely those issues that had not been adequately answered by the previous committees and task forces. One of the goals that the Task force on the Artifact in Library Collections has been to propose mechanisms for collaboration in the task of articulating preservation strategies that are realistic and sustainable, balancing the needs of present users with those of the future. As chair of the task force, I deemed that before we could hope to propose collaborative mechanisms for balancing use and preservation, we would have to propose a use-based definition of the artifact that would be, from the outset, correlated with the requirements and realities of preservation. Space constraints do not permit me to present the full extent of our thinking and recommendations here. I did want to share with you, however, the basis for our thinking. I hope to have succeeded.

---

[8] The resulting report, *Preserving Research Collections*, was published in 1999. (Reed-Scott 1999).